



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Compressive and Noncompressive Power Spectral Density Estimation from Periodic Nonuniform Samples

Citation for published version:

Lexa, MA, E. Davies, M & S. Thompson, J 2011 'Compressive and Noncompressive Power Spectral Density Estimation from Periodic Nonuniform Samples' ArXiv. <<http://arxiv.org/abs/1110.2722v2>>

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Early version, also known as pre-print

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Compressive and Noncompressive Power Spectral Density Estimation from Periodic Nonuniform Samples

Michael A. Lexa, Mike E. Davies and John S. Thompson
 {michael.lexa,mike.davies,john.thompson}@ed.ac.uk

Institute of Digital Communications
 University of Edinburgh

12 Oct 2011, Revised 15 Apr 2012

Abstract

This paper presents a novel power spectral density estimation technique for band-limited, wide-sense stationary signals from sub-Nyquist sampled data. The technique employs multi-coset sampling and incorporates the advantages of compressed sensing (CS) when the power spectrum is sparse, but applies to sparse and nonsparse power spectra alike. The estimates are consistent piecewise constant approximations whose resolutions (width of the piecewise constant segments) are controlled by the periodicity of the multi-coset sampling. We show that compressive estimates exhibit better tradeoffs among the estimator's resolution, system complexity, and average sampling rate compared to their noncompressive counterparts. For suitable sampling patterns, noncompressive estimates are obtained as least squares solutions. Because of the non-negativity of power spectra, compressive estimates can be computed by seeking non-negative least squares solutions (provided appropriate sampling patterns exist) instead of using standard CS recovery algorithms. This flexibility suggests a reduction in computational overhead for systems estimating both sparse and nonsparse power spectra because one algorithm can be used to compute both compressive and noncompressive estimates.

1 Introduction

Compressed sensing (CS) is a data acquisition strategy that exploits the sparsity or compressibility of a signal [1–4]. Typically, a signal is defined to be *sparse* if its representation in an orthogonal basis contains only a few nonzero coefficients and is *compressible* if it can be well approximated by a sparse signal. In its simplest setting, CS proposes to sample a sparse discrete signal $\mathbf{x} \in \mathbb{R}^n$ by acquiring $m < n$ inner product measurements $y_l = \langle \mathbf{a}_l, \mathbf{x} \rangle$, $0 \leq l \leq m$, where \mathbf{a}_l are the rows of a $m \times n$ measurement matrix \mathbf{A} . In matrix notation, $\mathbf{y} = \mathbf{A}\mathbf{x}$. Remarkably, CS asserts that \mathbf{x} can be accurately (if not exactly) reconstructed from the measurements \mathbf{y} even when the linear system of equations is underdetermined, provided \mathbf{A} satisfies certain conditions. When CS is applied to the problem of sampling continuous-time signals, it has been shown that the undersampling of finite length vectors translates into sub-Nyquist sampling rates, where CS provides the theoretical justification and tools to reconstruct the signal despite the spectral aliasing that can occur [5–8].

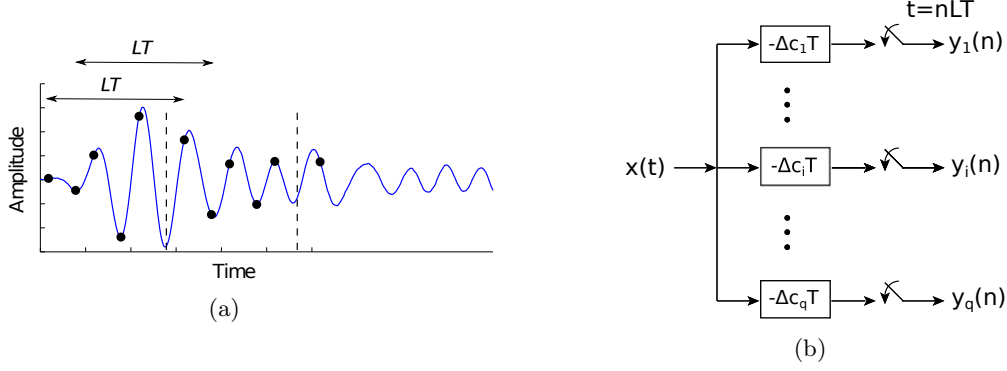


Figure 1: (a) An example of multi-coset sampling. Samples (dots) are acquired at nonuniformly spaced points over LT seconds with additional samples acquired at the same relative points in time over successive blocks of LT seconds. (b) Multi-coset sampler implemented as a multi-channel system.

In this paper, we study the problem of estimating the power spectral density (PSD) of a wide-sense stationary (WSS) random process within the context of a sub-Nyquist nonuniform periodic sampling strategy known as multi-coset (MC) sampling [9–15]. Because of its nonuniform periodic nature, MC sampling is closely related to CS,¹ and consequently, the proposed estimator easily incorporates and exploits CS theory. The estimator however applies broadly, and in particular, applies to cases where the PSD is not sparse.

For a fixed time interval T and for a suitable positive integer L , MC samplers sample an input signal $x(t)$ at the time instants $t = (nL + c_i)T$ for $1 \leq i \leq q$, $n \in \mathbb{Z}$, where the time offsets c_i are distinct, non-negative integers less than L ($0 \leq c_i < L$). The strategy is depicted in Fig. 1. The sampling times are known collectively as the multi-coset *sampling pattern* and the sets $\{(nL + c_i)T : n \in \mathbb{Z}, c_i \in \{0, \dots, L-1\}\}$ are cosets of $\{nT : n \in \mathbb{Z}\}$. Because T is fixed throughout the paper, we simply refer to $\{c_i\}$ as the multi-coset sampling pattern instead of the actual sampling times. MC samplers are parameterized by q , L , and $\{c_i\}$ and exhibit an average sampling rate of q/LT Hz. They are most easily implemented as multichannel systems as shown in Fig. 1 where channel i shifts $x(t)$ by $c_i T$ seconds and then samples uniformly at $1/LT$ Hz.

Here, we assume a multichannel MC sampler and propose a novel nonparametric PSD estimator based on the MC samples. The method estimates the average power within specific subbands of a WSS random process. Hence it produces piecewise constant estimates that are in contrast to those supported on a discrete frequency grid, e.g., the periodogram implemented using the discrete Fourier transform (DFT). Moreover, the estimator does not use passband filters to isolate subband signal components. Rather the estimator uses the spectral aliasing that occurs in each channel to underpin the formation of a linear system of equations whose solution is the PSD estimate of interest (see Section 2 for details).

We categorize the solutions to this linear system as either *compressive* or *noncompressive* depending upon whether CS is employed in the recovery of the estimate. Compressive estimates result from the application of CS theory and pertain to situations where the underlying power spectrum is sparse and the linear system of equations is underdetermined. Noncompressive

¹In fact, it can be formulated as a CS sampling strategy.

estimates pertain to cases where the linear system is overdetermined (or determined) with the power spectrum's sparsity being irrelevant. In both cases, the uniqueness of the recovered estimates depends on the choice of the sampling pattern $\{c_i\}$ (not all sampling patterns lead to unique solutions). Because of the special structure of the linear system, we show that in some cases compressive estimates can be computed using a non-negative least squares (NNLS) approach. This result is advantageous because in these cases it eliminates the necessity of using two separate algorithms to compute compressive and noncompressive estimates. (Note noncompressive estimates are normally computed as least squares solutions but can also be computed using NNLS; see Section 3.2 for details).

The proposed method also indicates the regions within the (L, q) parameter space that allow for compressive and noncompressive estimates, and it illustrates the associated tradeoffs among quantities related to these parameters such as the estimator's resolution or the number of MC channels. By comparing these tradeoffs one can gauge the benefit of applying CS theory in cases where the PSD is sparse. Perhaps surprisingly, the advantage of a compressive versus a noncompressive estimate is not simply an improved sampling rate, but rather improved tradeoffs [16]. For example, for a fixed spectral resolution, a compressive estimate might achieve the same accuracy as a noncompressive estimate but with a smaller number of channels. In addition, we show that both compressive and noncompressive estimates are statistically consistent and show that the noncompressive estimate is asymptotically efficient.

1.1 Related work

There is a growing literature in applying CS and other sub-Nyquist techniques to power spectrum estimation, most notably in the field of cognitive radios where there is a need to find underutilized bandwidth in a crowded spectrum [17–22]. Most pertinent to the present discussion, however, is the work of Ariananda, Leus and Tian [18–20]. These papers also address the problem of PSD estimation from sub-Nyquist samples specifically incorporating CS-based sampling systems (including MC sampling). Notwithstanding similar goals and approaches, the estimator proposed here differs in two significant aspects. First, their method estimates the cross-correlations of the MC sample sequences and then recovers an estimate of the autocorrelation sequence of the input random process (from which they form a periodogram estimate of the PSD). In contrast, the method described here directly recovers a piecewise constant estimate from the MC cross-correlation estimates. The main difference being that Ariananda et al. estimate *samples* of the PSD whereas we estimate the *average power within subbands*. An implication of this difference is that the method of Ariananda et al. is more computationally expensive: their method requires the formation of a large $q^2 \times L^2$ matrix in the construction of their $q^2 \times (2L - 1)$ measurement matrix; in contrast, we do not require the construction of this large matrix for our smaller $q(q - 1)/2 + 1 \times L$ measurement matrix. Second, we treat the compressive case and address the statistical nature of our estimator (specifically taking into account the finite availability of data), while Ariananda et al. do not.

1.2 Main contribution

In short, the paper's main contribution is the development of a new nonparametric, statistically consistent, piecewise constant PSD estimator based on sub-Nyquist MC samples. When the underlying PSD is sparse, it seamlessly incorporates CS theory, letting the resulting compressive

estimator exhibit better parametric tradeoffs compared to their noncompressive counterparts. Moreover, when a suitable sampling pattern exists, a NNLS approach can be used to compute compressive estimates, suggesting a possible savings in computational overhead for systems estimating sparse and nonsparse power spectra.

2 A PSD approximation from multi-coset samples

In this section, we derive a piecewise constant PSD approximation assuming full knowledge of the MC sequences. This approximation serves as the motivation and model for the noncompressive and compressive estimators derived in Sections 3. In Section 4, we present three examples illustrating the character and properties of the estimators.

2.1 PSD approximation

Let $x(t)$ be a real, zero-mean, WSS random process with autocorrelation function $r_{xx}(\tau) \triangleq \mathbb{E} x(t_1)x(t_2)$, $\tau = t_1 - t_2$. The Fourier transform (FT) $P_{xx}(\omega)$ of $r_{xx}(\tau)$ is the PSD of $x(t)$ [23].

$$r_{xx}(\tau) \xleftrightarrow{\text{FT}} P_{xx}(\omega)$$

The PSD quantifies the power in any given spectral band of $x(t)$ and is a second order description of the random process. Because $r_{xx}(\tau)$ is real and symmetric, $P_{xx}(\omega)$ is also real and symmetric. Throughout the paper, we assume $x(t)$ is a band-limited process, meaning that $P_{xx}(\omega)$ vanishes for $|\omega| > \pi W$ rad/s [23, p. 376]. For the MC sampling, we set $T = 1/W$ thereby referencing the average system sampling rate qW/L to the Nyquist rate W .

To estimate $P_{xx}(\omega)$ using MC samples, we examine the cross-correlations of the output sequences $y_i(n) \triangleq x(n\frac{L}{W} + \frac{c_i}{W})$, $n \in \mathbb{Z}$. Let a and b denote two channels indices, then

$$\begin{aligned} r_{y_a y_b}(n, m) &= \mathbb{E} y_a(n) y_b(m) \\ &= \mathbb{E} x(n\frac{L}{W} + \frac{c_a}{W}) x(m\frac{L}{W} + \frac{c_b}{W}) \\ &= r_{xx}(\frac{L}{W}(n - m) + \frac{1}{W}(c_a - c_b)) \\ &= r_{xx}(\frac{L}{W}k + \frac{1}{W}(c_a - c_b)) \\ &= r_{y_a y_b}(k) \end{aligned} \tag{1}$$

where the third step follows from the wide-sense stationarity of $x(t)$ and the substitution $k = n - m$. From (1), we see that the cross-correlation function $r_{y_a y_b}(k)$ is equivalent to shifting the auto-correlation function $r_{xx}(\tau)$ by $\frac{1}{W}(c_a - c_b)$ and then uniformly sampling it at W/L Hz. These operations imply the following discrete time Fourier transform (DTFT) pair for $k \in \mathbb{Z}$ and $\omega \in \mathbb{R}$,

$$\begin{aligned} r_{y_a y_b}(k) &= r_{xx}(\frac{L}{W}k + \frac{1}{W}(c_a - c_b)) \\ &\quad \updownarrow \text{DTFT} \\ \frac{W}{L} \sum_{m=\lceil \frac{L}{2}(\frac{\omega}{\pi W} - 1) \rceil}^{\lfloor \frac{L}{2}(\frac{\omega}{\pi W} + 1) \rfloor} P_{xx}(\omega - 2\pi\frac{W}{L}m) e^{j\frac{1}{W}(c_a - c_b)(\omega - 2\pi\frac{W}{L}m)} & \end{aligned} \tag{2}$$

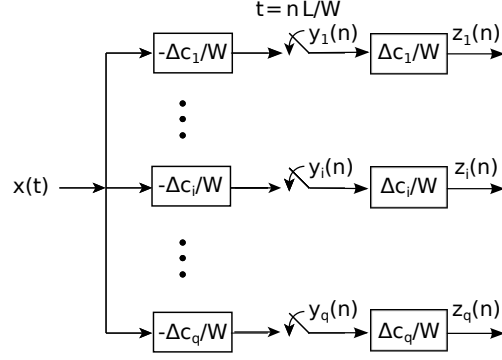


Figure 2: Block diagram of the multi-coset sampling system that shows the temporal shifting before and after sampling.

where the relation follows from the shifting and sampling properties of the DTFT [24, p. 98, 117]. The summation limits are finite for a given ω because $x(t)$ is assumed band-limited. (Here we also assume L is even.)

The phase shift $e^{j\frac{1}{W}(c_a-c_b)\omega}$ in (2) arises from the time shifts introduced in the individual channels. For our purposes, however, this phase shift needs to be removed. This can be accomplished by shifting the sequences $y_i(n)$ by an amount equal to, but opposite, the initial delays (see Fig. 2). Mathematically, this step can be expressed as

$$z_i(n) \triangleq (y_i \star h_i)(n), \quad n \in \mathbb{Z}, \quad i = 1, \dots, q \quad (3)$$

where \star denotes convolution and $h_i(k) = \text{sinc}(\pi(k - c_i/W))$, $i = 1, \dots, q$, are the impulse responses of ideal fractional delay filters [25]. These filters are “fractional” in the sense that the time shifts are a fraction of the sampling period L/W . For simplicity, we assume ideal fractional delay filters throughout the paper. We note however that there is a large literature concerning the design and analysis of digital fractional delay filters (see [25] and the references therein) ranging from highly accurate and computationally expensive filters to less accurate and computationally inexpensive ones. The choice of design is largely application dependent and a detailed analysis of the impact of imperfect fractional delay filters and is beyond the scope of this paper.

By examining the cross-correlation functions $r_{z_a z_b}(k)$, we see that the fractional delays indeed delay $r_{y_a y_b}(k)$ by $(c_a - c_b)/W$ seconds.

$$\begin{aligned} r_{z_a z_b}(k) &= \mathbb{E} z_a(k+n) z_b(n) \\ &= \sum_m \sum_l h_a(m) h_b(l) r_{y_a y_b}(k - m + l) \\ &= \sum_{\alpha} \left[\sum_m h_a(m) r_{y_a y_b}(\alpha - m) \right] h_b(\alpha - k) \end{aligned} \quad (4)$$

$$\begin{aligned} &= r_{y_a y_b}(k) \star h_a(k) \star h_b(-k) \\ &= r_{y_a y_b}(k) \star h_{a-b}(k) \end{aligned} \quad (5)$$

where the second step follows from (3), $\alpha = k + l$, and $h_{a-b}(k)$ denotes the impulse of an ideal fractional delay filter with delay $(c_a - c_b)/W$ seconds.

Equation (5) also implies that the DTFT of $r_{z_a z_b}(k)$ is the DTFT of $r_{y_a y_b}(k)$ multiplied by $e^{-j\frac{1}{W}(c_a - c_b)\omega}$. Thus, by multiplying the frequency domain component in (2) by $e^{-j\frac{1}{W}(c_a - c_b)\omega}$, we obtain the following *inverse* DTFT relation

$$r_{z_a z_b}(k) = \frac{1}{2\pi} \sum_{m=-L/2+1}^{L/2} e^{-j\frac{2\pi}{L}(c_a - c_b)m} \int_{-\pi W/L}^{\pi W/L} P_{xx}(\omega - 2\pi\frac{W}{L}m) e^{jk\frac{L}{W}\omega} d\omega. \quad (6)$$

Evaluating both sides of this equation at $k = 0$ produces a linear system of equations over the set of all pairs (a, b) of channels:

$$r_{z_a z_b}(0) = \sum_{m=-L/2+1}^{L/2} e^{-j\frac{2\pi}{L}(c_a - c_b)m} \underbrace{\int_{-\pi W/L}^{\pi W/L} P_{xx}(\omega - 2\pi\frac{W}{L}m) \frac{d\omega}{2\pi}}_{\triangleq P_{xx}(m)}. \quad (7)$$

This linear system is the basis for our estimator. For a given m , $P_{xx}(m)$ equals the average power of the process $x(t)$ in the spectral subband $[(2m-1)\pi W/L, (2m+1)\pi W/L]$. Hence the set $\{\frac{L}{W}P_{xx}(m)\}$ forms a piecewise constant approximation to $P_{xx}(\omega)$. Its resolution equals the width of the piecewise constant segments (W/L Hz) and is inversely proportional to L , the parameter determining the period of the nonuniform samples. Larger L implies finer resolution; smaller L implies coarser resolution. The left hand side of (7) is a weighted sum of the total average power of $x(t)$. In Section 3, we propose a statistical estimator based on this approximation that uses only a finite number of multi-coset samples. Like the approximation $\{\frac{L}{W}P_{xx}(m)\}$, the estimator is piecewise constant and thus does not estimate PSD amplitudes at specific frequencies like standard parametric techniques or periodogram estimates implemented using the DFT.

Letting i index the $\binom{q}{2} + 1$ combinations² of pairs (a, b) , we can express (7) in matrix notation,

$$\mathbf{u} = \mathbf{\Psi} \mathbf{v} \quad (8)$$

where $\mathbf{u} \in \mathbb{R}^{q(q-1)/2+1}$, $\mathbf{v} \in \mathbb{R}^L$ and

$$\begin{aligned} \mathbf{u} &= [u_0, \dots, u_{q(q-1)/2}]^T \text{ (T denotes transpose)} \\ u_i &= r_{z_a z_b}(0) \\ \Psi_{i,l} &= e^{-j\frac{2\pi}{L}(c_a - c_b)_i m_l} \\ \mathbf{v} &= [v_0, \dots, v_{L-1}]^T \\ v_l &= P_{xx}(m_l) \end{aligned} \quad (9)$$

²In actuality, the total number of pairs is $\binom{q}{2} + q$, but each of the q cases where $a = b$ contributes a row of ones to $\mathbf{\Psi}$. Thus to avoid the row dependency, only a single row of ones is included in $\mathbf{\Psi}$.

for $i = 0, \dots, q(q-1)/2$, $l = 0, \dots, L-1$, and $m_l = -L/2 + 1 + l$. The number of equations in this linear system can be doubled by separating Ψ and \mathbf{u} into their real and imagery parts:

$$\underbrace{\begin{bmatrix} \text{Re}(\mathbf{u}) \\ \text{Im}(\mathbf{u}) \end{bmatrix}}_{\triangleq \tilde{\mathbf{u}}} = \underbrace{\begin{bmatrix} \text{Re}(\Psi) \\ \text{Im}(\Psi) \end{bmatrix}}_{\triangleq \tilde{\Psi}} \mathbf{v} \quad (10)$$

where $\tilde{\mathbf{u}}$ and $\tilde{\Psi}$ have dimensions $q(q-1) + 1 \times 1$ and $q(q-1) + 1 \times L$, respectively.

To calculate the PSD approximation, one would compute the set of cross-correlation values $\{r_{z_a z_b}(0) = \mathbb{E} z_a(n) z_b(n)\}$ for all pairs (a, b) and solve (10) for \mathbf{v} . Here we briefly consider two types of solutions, deferring a more detailed discussion until after the estimator is presented in Section 3 (see in particular Sections 3.2 and 3.3). First, we consider least squares solutions when $\tilde{\Psi}$ is full rank (rank L) and (10) is overdetermined. Such solutions are referred to as *noncompressive*. Second, we consider CS solutions when (10) is underdetermined and \mathbf{v} is sparse in the sense that only a few of its entries are nonzero. These solutions are termed *compressive*.

The existence of a unique least squares solution depends on the (column) rank of $\tilde{\Psi}$ which in turn depends on the sampling pattern $\{c_i\}$. In particular, a sampling pattern must be chosen such that $\tilde{\Psi}$ has rank L (implying that $q(q-1) + 1 \geq L$, or equivalently, that the number of rows of $\tilde{\Psi}$ is greater than or equal to the number of columns). In Section 3.2, we discuss two methods for constructing sampling patterns that produce full rank $\tilde{\Psi}$, but given such a sampling pattern, the unique least squares solution is

$$\begin{aligned} \mathbf{v}_{NC} &\triangleq \arg \min_{\alpha} \|\tilde{\Psi} \alpha - \tilde{\mathbf{u}}\|_2^2 \\ &= (\tilde{\Psi}^T \tilde{\Psi})^{-1} \tilde{\Psi}^T \tilde{\mathbf{u}} \end{aligned} \quad (11)$$

where $\|\cdot\|_2$ denotes the ℓ_2 norm, $(\tilde{\Psi}^T \tilde{\Psi})^{-1} \tilde{\Psi}^T$ equals the pseudoinverse of $\tilde{\Psi}$, and the subscript NC indicates that the approximation is noncompressive. If $q(q-1)/2 + 1 = L$ and $\tilde{\Psi}$ is nonsingular, we simply have $\mathbf{v}_{NC} = \tilde{\Psi}^{-1} \tilde{\mathbf{u}}$.

Now, suppose $P_{xx}(\omega)$ is spectrally sparse, i.e., suppose its support has Lebesgue measure that is small relative to the overall bandwidth. In addition, suppose $q(q-1) + 1 < L$ such that (10) is an underdetermined linear system of equations. Then from a CS perspective, \mathbf{v} is the sparse vector of interest that is to be recovered from the linear measurements $\tilde{\mathbf{u}}$, and $\tilde{\Psi}$ is a (subsampled Fourier) measurement matrix. The vector \mathbf{v} is sparse because its entries represent the average power of $x(t)$ in the subbands $[(2m-1)\pi W/L, (2m+1)\pi W/L]$ and $P_{xx}(\omega)$ is presumed to be sparse. In this situation, a host of CS recovery algorithms may be used to compute compressive approximations \mathbf{v}_C provided an appropriate sampling pattern and a sufficient number of measurements for a given level of sparsity (see e.g., [26–28]). For example, for a suitable sampling pattern, the convex optimization problem

$$\min_{\alpha} \|\alpha\|_1 \text{ subject to } \tilde{\mathbf{u}} = \tilde{\Psi} \alpha, \quad (12)$$

would yield a compressive approximation if $\mathcal{O}(S \log^4 L)$ measurements were acquired [29], where $\|\cdot\|_1$ denotes the ℓ_1 norm and S denotes the number of nonzero entries in \mathbf{v} . In Section 3.3, we further exploit the structure of this linear system and advocate an algorithm that does not require an explicit sparsity constraint in the recovery procedure.

3 Proposed multi-coset PSD estimator

The above PSD approximation is predicated on full knowledge of the output sequences $y_i(n)$, or equivalently, on having access to an infinite amount of data (an infinite number of multi-coset samples). In this section, we derive and characterize an estimator based on (8) that explicitly accounts for the finite availability of data.

3.1 PSD estimation

Again, let $x(t)$ be a real, band-limited, zero-mean WSS random process with PSD $P_{xx}(\omega)$. We observe $x(t)$ over a finite duration interval and hence model the observed signal $x^D(t)$ as a windowed version of $x(t)$:

$$x^D(t) \triangleq x(t)w_R(t) \quad (13)$$

where

$$w_R(t) = \begin{cases} 1 & 0 \leq t \leq D \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

The MC samples of the windowed process $x^D(t)$ are

$$\begin{aligned} y_i^N(n) &\triangleq x^D\left(\frac{1}{W}(nL + c_i)\right) \\ &= x\left(\frac{1}{W}(nL + c_i)\right)w_R\left(\frac{1}{W}(nL + c_i)\right), \end{aligned} \quad (15)$$

for $n = 0, \dots, N-1$ and $i = 1, \dots, q$, where $N = DW/L$ equals the number of samples acquired per channel during the D second observation interval. (Here N is assumed to be an integer.) The fractionally delayed samples are (c.f. (3))

$$z_i^N(n) \triangleq (y_i^N \star h_i)(n) \quad (16)$$

where again N signifies the number of samples³.

Let a and b denote the indices of two channels. Then, given $z_a^N(n)$ and $z_b^N(n)$, the cross-correlation function $r_{z_a z_b}(k)$ can be estimated by the sample correlation

$$\hat{r}_{z_a z_b}^N(k) = \frac{1}{N} \sum_{n=0}^{N-|k|-1} z_a^N(k+n)z_b^N(n), \quad (17)$$

for $0 \leq |k| \leq N-1$. Mimicking the standard derivation of a periodogram (see e.g. [30, p.321]), we find the DTFT of $\hat{r}_{z_a z_b}^N(k)$ to be

$$\hat{r}_{z_a z_b}^N(k) \xrightarrow{\text{DTFT}} \frac{1}{N} \left[Z_a^N(e^{j\omega \frac{L}{W}}) \right] \left[Z_b^N(e^{j\omega \frac{L}{W}}) \right]^* \quad (18)$$

where $Z_a^N(e^{j\omega \frac{L}{W}})$ and $Z_b^N(e^{j\omega \frac{L}{W}})$ are the DTFT of $z_a^N(n)$ and $z_b^N(n)$ respectively. The transform $Z_a^N(e^{j\omega \frac{L}{W}})$ may be further decomposed in terms of $Y_a^N(e^{j\omega \frac{L}{W}})$ and $X^D(\omega)$ where

$$\begin{aligned} x^D(t) &\xrightarrow{\text{FT}} X^D(\omega) \\ y_a^N(n) &\xrightarrow{\text{DTFT}} Y_a^N(e^{j\omega \frac{L}{W}}). \end{aligned} \quad (19)$$

³A perfect fractional delay filter has an infinite impulse response $h_i(n) = \text{sinc}(\pi(n - c_i/W))$ and thus the sequence $z_i^N(n)$ has, strictly speaking, infinite duration. However, in (16) we only consider N output samples $z_i^N(n)$, $n = 0, 1, \dots, N-1$. We do not include another windowing operation because $z_i^N(n)$ will be finite for any practical, realizable filter.

Specifically, for $\omega \in \mathbb{R}$, we have from (15) and (16) that

$$\begin{aligned} Z_a^N(e^{j\omega \frac{L}{W}}) &= Y_a^N(e^{j\omega \frac{L}{W}})e^{-j\frac{c_a}{W}\omega} \\ &= \frac{W}{L} \sum_{m=-\infty}^{\infty} X^D(\omega - 2\pi \frac{W}{L}m)e^{-j\frac{2\pi}{L}c_a m}. \end{aligned} \quad (20)$$

Because $x^D(t)$ is time limited, it cannot, strictly speaking, be band-limited. However, for sufficiently long windows it is reasonable to assume that the vast majority of the energy of $x^D(t)$ lies within the band $(-\pi W, \pi W]$, i.e., it is reasonable to assume that with sufficiently long windows $x^D(t)$ is essentially band-limited to the same band as $x(t)$. With this assumption, one period of $Z_a^N(e^{j\omega \frac{L}{W}})$ is well approximated by the finite summation,

$$Z_a^N(e^{j\omega \frac{L}{W}}) = \frac{W}{L} \sum_{m=-\frac{L}{2}+1}^{\frac{L}{2}} X^D(\omega - 2\pi \frac{W}{L}m)e^{-j\frac{2\pi}{L}c_a m}. \quad (21)$$

Substituting (21) for $Z_a^N(e^{j\omega \frac{L}{W}})$ in (18) (along with a corresponding expression for $Z_b^N(e^{j\omega \frac{L}{W}})$) and evaluating the inverse DTFT at zero yields

$$\begin{aligned} \hat{r}_{z_a z_b}^N(0) &= \sum_m \sum_n e^{-j\frac{2\pi}{L}(c_a m - c_b n)} \\ &\quad \int_{-\pi W/L}^{\pi W/L} \frac{1}{D} X_m^D(\omega) [X_n^D(\omega)]^* \frac{d\omega}{2\pi}, \end{aligned} \quad (22)$$

where $X_m^D(\omega)$ is shorthand notation for $X^D(\omega - 2\pi \frac{W}{L}m)$.

We observe that for $m = n$, the integrands, $\frac{1}{D} |X_m^D(\omega)|^2$, $m = -L/2 + 1, \dots, L/2$, are periodogram estimates of L disjoint subbands of $P_{xx}(\omega)$, and the corresponding integrals

$$\hat{P}_{xx}(m) \triangleq \int_{-\pi W/L}^{\pi W/L} \frac{1}{D} |X_m^D(\omega)|^2 \frac{d\omega}{2\pi} \quad (23)$$

are estimates of the power within these subbands. The set $\{\frac{L}{W} \hat{P}_{xx}(m)\}$ thus constitutes a piecewise constant estimate of $P_{xx}(\omega)$ at resolution W/L .

In contrast, the cross terms in (22) (i.e., terms for which $m \neq n$) asymptotically approach zero as D , or equivalently N , grows large. To see this, rewrite the integrand in (22) as

$$\begin{aligned} &\frac{1}{D} X_m^D(\omega) [X_n^D(\omega)]^* \\ &= \frac{1}{D} X^D(\omega - 2\pi \frac{W}{L}m) [X^D(\omega - 2\pi \frac{W}{L}n)]^* \\ &= \int_{-\infty}^{\infty} W_R(\alpha) X(\omega - 2\pi \frac{W}{L}m - \alpha) \frac{d\alpha}{\sqrt{D}} \\ &\quad \times \int_{-\infty}^{\infty} W_R^*(\beta) X^*(\omega - 2\pi \frac{W}{L}n - \beta) \frac{d\beta}{\sqrt{D}}, \end{aligned} \quad (24)$$

where $W_R(\omega) = D \operatorname{sinc}(\frac{D}{2}\omega) e^{-j\frac{D}{2}\omega}$ is the FT of the rectangular window $w_R(t)$, $X(\omega)$ is the FT of $x(t)$ appropriately defined (see [23, p. 416]), and

$$\operatorname{sinc}(\alpha) = \begin{cases} 1, & \text{if } \alpha = 0 \\ \sin(\alpha)/\alpha, & \text{otherwise} \end{cases}.$$

As $D \rightarrow \infty$, $W_R(\omega)$ approaches a delta function [31, p. 280]; however because of the factors $1/\sqrt{D}$, the integrals approach zero as $D \rightarrow \infty$ for $m \neq n$.

Because of this fact, we choose to approximate $\hat{r}_{z_a z_b}^N(0)$ using only the terms $m = n$ in (22)

$$\hat{r}_{z_a z_b}^N(0) \approx \sum_m e^{-j\frac{2\pi}{L}(c_a - c_b)m} \hat{P}_{xx}(m). \quad (25)$$

Letting i index all pairs (a, b) , we arrive at an empirical counterpart to (8)

$$\hat{\mathbf{u}} = \Psi \hat{\mathbf{v}} \quad (26)$$

where

$$\begin{aligned} \hat{\mathbf{u}} &= [\hat{u}_0, \dots, \hat{u}_{q(q-1)/2}]^T \\ \hat{u}_i &= \hat{r}_{z_a z_b}^N(0) \\ \Psi_{i,l} &= e^{-j\frac{2\pi}{L}(c_a - c_b)_i m_l} \\ \hat{\mathbf{v}} &= [\hat{v}_0, \dots, \hat{v}_{L-1}]^T \\ \hat{v}_l &\approx \hat{P}_{xx}(m_l) \end{aligned} \quad (27)$$

for $i = 0, \dots, q(q-1)/2$, $l = 0, \dots, L-1$, and $m_l = -L/2 + 1 + l$. Similar to (10), we can double the number of equations by separating the real and imagery parts of Ψ and $\hat{\mathbf{u}}$:

$$\tilde{\mathbf{u}} = \tilde{\Psi} \hat{\mathbf{v}} \quad (28)$$

where $\tilde{\mathbf{u}} = [\operatorname{Re}(\hat{\mathbf{u}}) \operatorname{Im}(\hat{\mathbf{u}})]^T$ and $\tilde{\Psi} \in \mathbb{R}^{q(q-1)+1 \times L}$ is defined in (10). In this equation, $\hat{\mathbf{v}}$ is the PSD estimator we wish to solve for.

3.2 Noncompressive solutions (overdetermined case)

A unique least squares solution to (28) exists if $\tilde{\Psi}$ has full column rank, i.e., $\operatorname{rank}(\tilde{\Psi}) = L$. Full rank implies $q(q-1)+1 \geq L$, or that the system is overdetermined (here we treat a determined system as a special case of the overdetermined one). In this case, a noncompressive estimate can be computed using the pseudoinverse

$$\hat{\mathbf{v}}_{NC} \triangleq (\tilde{\Psi}^T \tilde{\Psi})^{-1} \tilde{\Psi}^T \tilde{\mathbf{u}}, \quad (29)$$

where $(\tilde{\Psi}^T \tilde{\Psi})^{-1} \tilde{\Psi}^T$ is the pseudoinverse of $\tilde{\Psi}$. A solution of this type is *noncompressive* because it does not rely on CS theory, and in particular, it does not require $\hat{\mathbf{v}}$ to be sparse.

For a fixed L , a full rank $\tilde{\Psi}$ requires (i) choosing q large enough such that $q(q-1)+1 \geq L$ and (ii) choosing a sampling pattern such that $\operatorname{rank}(\tilde{\Psi}) = L$. The first condition is required since $q(q-1)+1 < L$ implies $\operatorname{rank}(\tilde{\Psi}) < q(q-1)+1$, regardless of the sampling pattern. The second condition is necessary since there exists sampling patterns that lead to $\tilde{\Psi}$ having

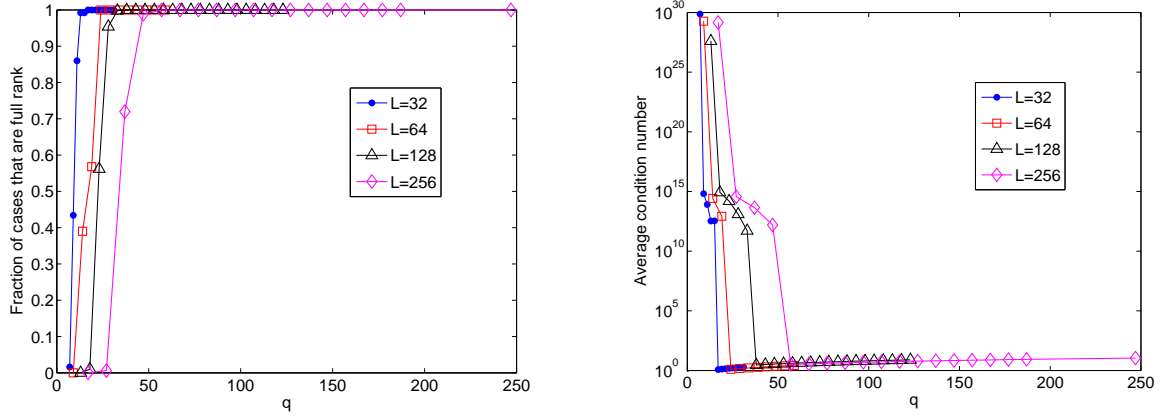


Figure 3: Four scenarios showing that random sampling patterns can produce well-conditioned, full rank $\tilde{\Psi}$ after some threshold point in the noncompressive case. Left: Fraction of times random patterns produce full rank $\tilde{\Psi}$ for different (L, q) values. Right: Corresponding condition numbers.

a non-empty nullspace even if $q(q-1)+1 \geq L$, meaning that the least squares solution is no longer unique. Section 3.4 discusses the tradeoff between q and L imposed by the first condition assuming the second condition is met. Here we address the second condition.

Fig. 3 offers empirical evidence suggesting that generating a sampling pattern uniformly at random from $\{0, \dots, L-1\}$ can be a simple and effective way of constructing a full rank $\tilde{\Psi}$, provided the ratio of its dimensions $\frac{L}{q(q-1)+1}$ is small enough. The plots show the results from four trials corresponding to four different values of L . For each trial, L was fixed and q was allowed to range from its minimum value—the smallest integer value satisfying $q(q-1)+1 > L$ —to its maximum value $L-1$. For each (L, q) pair, 500 sampling patterns (of length q) were generated and the fraction of those that produced full rank $\tilde{\Psi}$ are recorded in the left hand plot. The right hand plot displays the average condition number of $\tilde{\Psi}$ in each case. Both plots suggest that after some threshold point random sampling patterns consistently produce well-conditioned and full rank matrices for the noncompressive problem. For the four cases shown, the threshold occurs when the ratio $\frac{L}{q(q-1)+1}$ roughly equals 0.12, but a formal treatment of this phenomenon is beyond the scope of this paper.

For larger $\frac{L}{q(q-1)+1}$ ratios (still falling within the noncompressive problem), full rank $\tilde{\Psi}$ can sometimes be constructed by choosing the sampling pattern to be a *Golomb ruler* [32]. Golomb rulers are integer sets whose difference sets contain unique elements. The cardinality of the integer set is the ruler's *order* and the largest difference between two integers is its *length*. The idea is that the integers represent marks on an imaginary ruler and the elements of the difference set are all the lengths that can be measured by the ruler. The goal in the study of Golomb rulers is to find the minimum length rulers for a given order. In the present context, however, we are only interested in using already tabulated rulers. For example, for $L = 64$ and $q = 10$, one can take the minimum-length order-10 ruler to be the sampling pattern

$$\{c_i\}_{i=1}^{10} = \{1, 2, 7, 11, 24, 27, 35, 42, 54, 56\}.$$

Note that the ruler's length, 55, is less than $L = 64$ as it should be for MC sampling. This pattern yields a full rank $\tilde{\Psi}$ with condition number 1.4 whereas in Fig. 3 random patterns

never produced a full rank matrix out of 500 test cases. The implication of this improvement is somewhat limited however because there are relatively few known (up to order 26) minimum length Golomb rulers. For instance with $L = 1024$, the minimum q value is 33, but there are no known minimum length Golomb rulers with orders greater than 26. When applicable, Golomb ruler sampling patterns can be an effective means to obtain a full rank Ψ when the ratio $\frac{L}{q(q-1)+1}$ is near the threshold value. Note also that Ariananda et al. advocate the use of Golomb rulers in their study of multi-coset PSD estimation, but they used it to ensure the rank of a binary matrix [20].

To summarize the foregoing, one computes a noncompressive estimate by fractionally delaying each channel of MC samples $y_i^N(n), i = 1, \dots, q$, computing the cross-correlation values $\{\hat{r}_{z_a z_b}^N(0)\}$, and solving (26) via (29).

3.3 Compressive solutions (underdetermined and sparse case)

When (28) is underdetermined and the sampling pattern is chosen uniformly at random from $\{0, \dots, L-1\}$, $\tilde{\Psi}$ is a randomly subsampled DFT matrix which is a well-known type of CS measurement matrix. Thus a number of CS recovery algorithms (e.g., [26–28]) will solve (28), provided $\hat{\mathbf{v}}$ is sparse and the number of CS measurements (length of $\hat{\mathbf{u}}$) is sufficient. However, because (28) has special structure, the usual CS recovery algorithms may be replaced by a non-negative least squares approach. This claim rests on the following result of Donoho and Tanner [33] (see also [34]):

Corollary 1 ([33], Corollary 1.4 and 1.5). *Assume d is even and for $n = 0, 1, 2, \dots, L-1$, let $\mathbf{A} \in \mathbb{R}^{d \times L}$ be the partial DFT matrix ($d < L$)*

$$\mathbf{A}_{k,n} = \begin{cases} \cos\left(\frac{\pi kn}{L}\right) & k = 0, 2, 4, \dots, d-2 \\ \sin\left(\frac{\pi(k+1)n}{L}\right) & k = 1, 3, 5, \dots, d-1 \end{cases}$$

Let \mathbf{x} be a non-negative s -sparse vector, i.e., let \mathbf{x} have at most s nonzero positive entries. Then given the linear measurements $\mathbf{b} = \mathbf{A}\mathbf{x}$, any approach that solves the linear system and maintains nonnegativity will correctly recover the s -sparse solution \mathbf{x} , provided $d \geq 2s$.

The corollary says that when \mathbf{x} is s -sparse and non-negative, one needs only $2s$ Fourier measurements to uniquely recover it. Moreover, the theorem states that \mathbf{x} can be recovered using *any* algorithm that produces a nonnegative solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$. The fact that a L dimensional s -sparse vector \mathbf{x} can be recovered from only $d < L$ Fourier measurements is an archetypical CS result. However, in contrast to typical CS recovery strategies, signal recovery in this case does not require explicit use of a sparsity constraint like ℓ_1 minimization (see for example (12)). The reason is that this result is a specific case of a more general phenomenon in the recovery of sparse nonnegative vectors for underdetermined linear system of equations. In essence, it is known that if the row span of a measurement matrix intersects the positive orthant and if the vector to be recovered is nonnegative then the solution to the underdetermined linear system is a singleton, provided the vector of interest is sparse enough [35, 36] (see also [37]). This finding is key since it implies that compressive PSD estimates can be obtained in a manner similar to noncompressive estimates, i.e., they can be obtained using NNLS (more about this point in a moment).

The application of Corollary 1 to (28) is straightforward. The required matrix \mathbf{A} is a permuted version of $\tilde{\Psi}$ provided the sampling pattern is chosen such that the set of differences $(c_a - c_b)$ form the consecutive sequence $0, 1, 2, \dots, s-1$. In particular, \mathbf{A} can be constructed from Ψ as

$$\mathbf{A}_{k,n} = \begin{cases} \text{Re}\{\Psi_{k/2,n}\} & k = 0, 2, 4, \dots \\ \text{Im}\{-\Psi_{(k+1)/2,n}\} & k = 1, 3, 5, \dots \end{cases} \quad (30)$$

The value for q is chosen such that $q(q-1)+1 \geq 2s$, i.e., such that the number of measurements (length of $\tilde{\mathbf{u}}$) is greater than twice the sparsity.

One strategy to construct a sampling pattern that forms the difference set $\{0, 1, 2, \dots, s-1\}$ is to again use Golomb rulers. Most Golomb rulers do not produce a difference set that is a consequence sequence up to its length, but because differences do not repeat for Golomb rulers, they tend to furnish nearly consecutive sequences. For example, the order-7 minimum length ruler $\{1, 3, 4, 11, 17, 22, 26\}$ produces the difference set:

$$\begin{aligned} &\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, \\ &13, 14, 15, 16, 18, 19, 21, 22, 23, 25\} \end{aligned} \quad (31)$$

which nearly forms the consecutive sequence $0, \dots, 25$, missing only the elements 12, 17, and 20. In Section 4.2, we present an example that uses this sampling pattern to successfully recover a 16-sparse vector $\hat{\mathbf{v}}_C$. If an appropriate Golomb ruler (sampling pattern) does not exist for a given L and s , one can use other CS recovery approaches (e.g., ℓ_1 minimization, CoSaMP [27], or iterative hard thresholding [28]) with a randomly generated sampling pattern.

The non-negativity requirement of Corollary 1 for $\hat{\mathbf{v}}$ is met asymptotically as $N \rightarrow \infty$. This claim can be inferred from Section 3 where it is shown that the entries of $\hat{\mathbf{v}}$ approach integrated periodogram estimates, which are non-negative by construction (see (22) through (25)). For PSD estimation, the non-negativity requirement is natural since by definition power spectra are non-negative.

When Corollary 1 can be applied, we compute compressive estimates $\hat{\mathbf{v}}_C$ using NNLS:

$$\hat{\mathbf{v}}_C = \arg \min_{\alpha} \|\tilde{\Psi}\alpha - \tilde{\mathbf{u}}\|_2^2 \text{ subject to } \alpha \geq 0, \quad (32)$$

which is an optimization problem that can be solved via a linear program. Choosing a NNLS approach for compressive estimates is advantageous because noncompressive estimates can also be computed using a NNLS algorithm (and is perhaps preferred since $\hat{\mathbf{v}}_{NC}$ can be negative; see Fig. 8).

3.4 Tradeoffs in noncompressive case

Given a sampling pattern that produces a full column rank $\tilde{\Psi}$, the relation $q(q-1)+1 \geq L$ establishes a tradeoff between q and L (Fig. 4) and tradeoffs among other related quantities (Fig. 5). For example, it establishes a tradeoff between system complexity and the estimate's resolution, where here system complexity is simply taken to be the number of channels q . As the left hand plot in Fig. 5 shows, finer resolution (smaller W/L) comes at the price of higher system complexity. As a concrete example, consider a WSS signal $x(t)$ band-limited to 1 GHz and a desired resolution of 5 MHz, implying $L = 400$. A noncompressive solution would then require at least 21 channels; however, if a resolution of 15 MHz suffices, the number of channels could be reduced to 13 (see Fig. 5).

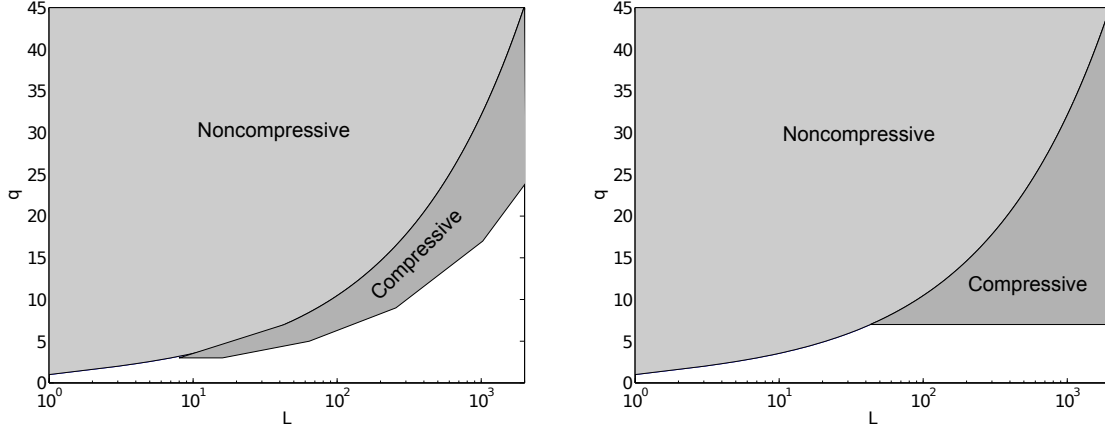


Figure 4: Tradeoffs between parameters q and L . The lightly shaded areas in both plots demarcate regions in the parameter space where noncompressive estimates exist given that $\tilde{\Psi}$ is full rank. The bounding curve is $q(q-1)+1=L$. The darker shaded areas are regions where compressive estimates exist and represent the possible gains in terms of improved parameter tradeoffs over the noncompressive case. The compressive regions derive from the specific scenario where $\hat{\mathbf{v}}$ is 16-sparse at resolution $L=128$.

If one has the freedom to choose q and L , the ratio q/L can be driven to zero even while maintaining the inequality $q(q-1)+1 \geq L$. This means that noncompressive estimates can theoretically be computed at arbitrarily low sampling rates and at finer and finer resolutions (see right hand plot of Fig. 5). Estimation at arbitrarily low sampling rates is a property shared with existing *alias-free* PSD estimators [38–42] that are based on random sampling where the set of sampling instances are (typically) realizations of a stochastic point process (e.g., a Poisson point process). Arbitrarily low sampling rates and arbitrarily fine resolutions, however, require arbitrarily high numbers of channels and arbitrarily long signal acquisition times. Thus in practical noncompressive situations, an appropriate balance needs to be struck between q and L that satisfies the requirements of a given application. Section 4.1 provides evidence of this tradeoff in a specific example.

3.5 Improved tradeoffs in compressive case

With sufficient sparsity, (28) can be uniquely solved even if it is underdetermined, i.e., even if $q(q-1)+1 < L$ [1, 3, 4]. This means that compressive estimates $\hat{\mathbf{v}}_C$ can be computed with better $q-L$ tradeoffs compared to noncompressive ones, i.e., for a given L , a compressive estimate can be recovered with a smaller q than what is possible for a noncompressive estimate, or conversely, for a given q , a compressive estimate can be recovered with a larger L .

The degree of the gain depends on the level of sparsity of $\hat{\mathbf{v}}$ at a specific resolution L . Let s denote the sparsity (number of nonzero elements) of $\hat{\mathbf{v}}$ at resolution L . Then according to Corollary 1, a compressive estimate can be computed with $2s$ or more Fourier measurements. Because the number of measurements in (28) equals $q(q-1)+1$ (length of $\tilde{\mathbf{u}}$), the need for $2s$ measurements sets a minimum q . Since compressive estimates pertain to underdetermined linear systems, the condition $q(q-1)+1 < L$ sets a maximum q . To apply Corollary 1, however, a sampling pattern of length q must exist that produces the set of differences $0, \dots, s-1$. If

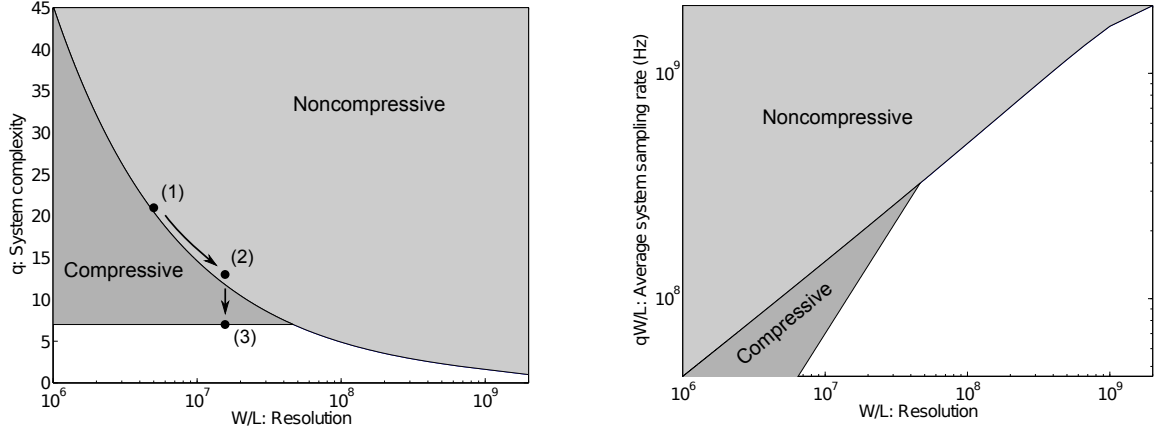


Figure 5: Tradeoffs among resolution, system complexity, and system sampling rate. Left: The progression from point (1) to point (2) shows that for a noncompressive estimator the reduction in the number of channels comes at a price of coarser resolution. However, for a 16-sparse vector $\hat{\mathbf{v}}$ at resolution 15 MHz, a compressive estimator can reduce the number of channels to 7—progression from (2) to (3). Right: Corresponding tradeoff between resolution and system sampling rate.

sampling pattern does not exist for the minimum q value, larger q values must be considered. Thus, for a compressive estimate, the value of q depends on the sparsity of $\hat{\mathbf{v}}$ (for a given L) and on the availability of an appropriate sampling pattern. We discuss specific examples in the next paragraph. Note that if an alternative CS recovery algorithm is used to solve (28) instead of applying Corollary 1, larger q values will be needed for a given s (and L) because these algorithms required larger numbers of measurements.

We illustrate the improved tradeoffs in two scenarios each of which are based on the presumption that $\hat{\mathbf{v}}$ is a 16-sparse vector at resolution $L = 128$ ($\hat{\mathbf{v}} \in \mathbb{R}^{128}$). The results are graphically shown in Fig. 4. As described above, the light shaded regions represent the (L, q) pairs for which noncompressive estimates can be computed⁴, and are bordered by the curve $q(q-1)+1=L$. The dark shaded regions in each panel represent the (L, q) pairs for which compressive estimates can be computed. The difference between the compressive regions on the left and right plots is that the region on the left presumes s changes with L and the one on the right presumes s is independent of L , i.e., it remains constant as L varies. In the left hand plot, we assume s doubles every time L doubles. Thus, in comparison to the initial presumption $s = 16$ at $L = 128$, we assume $s = 32$ at $L = 256$ which implies a minimum $q = 9$. This behavior models the situation where the active subbands of $x(t)$ at resolution $L = 128$ contain energy at every frequency within the subband, so as the subbands are subdivided, s increases. The scenario on the right reflects the situation where the active subbands only contain energy at one specific frequency, so s remains constant as L increases.

To judge the implications of these tradeoffs, we examine the associated tradeoffs among system complexity (q), resolution (W/L), and average system sampling rate (qW/L) in Fig. 5. As above, the compressive regions are drawn under the presumption that $s = 16$ for $L = 128$. The plots, however, are only for the case where s is independent of L . Intuitively, this

⁴Again, the ability to compute a noncompressive estimate for a valid (L, q) pair also depends on whether a suitable sampling pattern can be found.

represents a best case scenario in terms of “CS gain” over the noncompressive case (best possible improvements in q for a fixed L and in L for a fixed q). The left hand panel shows system complexity as a function of resolution W/L where $W = 2$ GHz. The right hand panel plots resolution versus system sampling rate. The plots show that for this particular scenario, a CS approach can have appreciable benefits when the underlying PSD is sparse. For example, in Section 3.4 we stated that the recovery of $\hat{\mathbf{v}}$ with a resolution of 15 MHz required a 13 channel MC sampler for a noncompressive estimator. However, for a 16-sparse vector $\hat{\mathbf{v}}$, a compressive estimate with the same resolution can be recovered with only $q = 7$ channels, a reduction of almost 1/2. This gain is depicted in Fig. 5 as the shift from point (2) to point (3). The same reduction factor can also be seen in average system sampling rate (right hand plot of Fig. 5).

These improved tradeoffs may or may not effect the error in the recovered estimate. If L is held fixed and $\hat{\mathbf{v}}$ is sparse, then as described above, q can be reduced to the smallest integer such that $q(q-1)+1 > 2s$ (according to Corollary 1). This reduction in q reduces the overall system sampling rate; however, the error in the recovered compressive estimate is *not* significantly different from the error in a corresponding noncompressive estimate because CS theory guarantees accurate (if not exact) recovery of $\hat{\mathbf{v}}$. Now, if q is held fixed and L is increased, the sampling rate of each channel decreases. Thus, if the observation interval is fixed, the correlation estimates \hat{u}_i become worse (in the sense that their error is larger) because less and less samples are used to compute \hat{u}_i . In this case, CS still allows the recovery of an estimate, but its quality is declining as L increases. To maintain accuracy, one must increase the observation interval and collect more samples per channel.

3.6 Consistency of the noncompressive estimator

To simplify notation, we examine the bias and variance of the least squares solution to $\hat{\mathbf{u}} = \Psi\hat{\mathbf{v}}$ instead of the solution of the expanded version $\hat{\tilde{\mathbf{u}}} = \tilde{\Psi}\hat{\mathbf{v}}$. The analysis is equivalent in either case. Assuming Ψ is full rank, the noncompressive solution is

$$\begin{aligned}\hat{\mathbf{v}}_{NC} &= (\Psi^H \Psi)^{-1} \Psi^H \hat{\mathbf{u}} \\ &= \Psi^\dagger \hat{\mathbf{u}},\end{aligned}$$

where H denotes Hermitian transpose and Ψ^\dagger is shorthand notation for the pseudoinverse.

Bias. The bias of $\hat{\mathbf{v}}_{NC}$ is the amount its expected value differs from $\mathbf{v}_{NC} = \Psi^\dagger \mathbf{u}$. Taking the expectation, we have $\mathbb{E} \hat{\mathbf{v}}_{NC} = \Psi^\dagger \mathbb{E} \hat{\mathbf{u}}$. One element of $\mathbb{E} \hat{\mathbf{u}}$ equals

$$\mathbb{E} \hat{u}_i = \mathbb{E} \hat{r}_{z_a z_b}^N(0) \tag{33}$$

$$= \frac{1}{N} \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h_a(m) h_b(l) \tag{34}$$

$$\begin{aligned}& \sum_{n=0}^{N-1} \mathbb{E} [y_a(n-m) y_b(n-l)] w_R(n-m) w_R(n-l) \\ &= \sum_m \sum_l h_a(m) h_b(l) r_{y_a y_b}(l-m) \\ & \quad \frac{1}{N} \sum_{n=0}^{N-1} w_R(n-m) w_R(n-l).\end{aligned} \tag{35}$$

where, as in (27), i is an index for the pair (a, b) , the second equality follows from (16) and (17), and $w_R(n)$ is a discrete rectangular window

$$w_R(n) = \begin{cases} 1 & 0 \leq n \leq N-1 \\ 0 & \text{otherwise.} \end{cases}$$

It is straightforward to show that

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} w_R(n-m) w_R(n-l) \\ = \begin{cases} 1 - \frac{|m-l|}{N} & |m-l| \leq N-1 \\ 0 & |m-l| > N-1. \end{cases} \end{aligned} \quad (36)$$

By substituting this expression back into (35), it is clear that $\hat{\mathbf{v}}_{NC}$ is a biased estimate for finite N ; however asymptotically, $\hat{\mathbf{v}}_{NC}$ is unbiased since

$$\mathbb{E} \hat{u}_i \stackrel{N \rightarrow \infty}{=} \sum_m \sum_l h_a(m) h_b(l) r_{y_a y_b}(l-m) \quad (37)$$

$$= r_{z_a z_b}(0) \quad (\text{from (4)}) \quad (38)$$

$$= u_i. \quad (39)$$

implies $\mathbb{E} \hat{\mathbf{v}}_{NC} = \mathbf{\Psi}^\dagger \mathbb{E} \hat{\mathbf{u}} \rightarrow \mathbf{\Psi}^\dagger \mathbf{u} = \mathbf{v}$ as $N \rightarrow \infty$.

Variance. The covariance matrix of $\hat{\mathbf{v}}_{NC}$ is

$$\begin{aligned} \mathbb{E} (\hat{\mathbf{v}}_{NC} - \mathbb{E} [\hat{\mathbf{v}}_{NC}]) (\hat{\mathbf{v}}_{NC} - \mathbb{E} [\hat{\mathbf{v}}_{NC}])^H \\ = \mathbf{\Psi}^\dagger \underbrace{\mathbb{E} (\hat{\mathbf{u}} - \mathbb{E} \hat{\mathbf{u}}) (\hat{\mathbf{u}} - \mathbb{E} \hat{\mathbf{u}})^H}_{\triangleq \mathbf{K}} [\mathbf{\Psi}^\dagger]^H \end{aligned} \quad (40)$$

where \mathbf{K} is the covariance matrix of $\hat{\mathbf{u}}$. The diagonal elements of (40) are the variances of the elements of $\hat{\mathbf{v}}_{NC}$:

$$\text{var}(\hat{v}_l) = \sum_i \sum_j \mathbf{\Psi}_{l,j}^\dagger \mathbf{K}_{i,j} \mathbf{\Psi}_{l,i}^\dagger, \quad l = 0, \dots, L-1 \quad (41)$$

where $\mathbf{K}_{i,j} = \text{covar}(\hat{u}_i, \hat{u}_j)$. Note that in (41) we have dropped the subscript NC in our notation for the elements of $\hat{\mathbf{v}}_{NC}$. To show that $\text{var}(\hat{v}_l) \rightarrow 0$ as $N \rightarrow \infty$, we proceed to upper bound (41) with a sequence in N that approaches zero as $N \rightarrow \infty$.

First, we note that the inequality $[\text{covar}(\hat{u}_i, \hat{u}_j)]^2 \leq \text{var}(\hat{u}_i) \text{var}(\hat{u}_j)$ [43, p. 336] implies

$$|\text{covar}(\hat{u}_i, \hat{u}_j)| \leq \sqrt{\text{var}(\hat{u}_i) \text{var}(\hat{u}_j)}. \quad (42)$$

Second, for the correlation estimate $\hat{u}_i = \hat{r}_{z_a z_b}^N(0)$ it is well-known [30, p. 320] that⁵

$$\text{var}(\hat{r}_{z_a z_b}^N(0)) \approx \frac{2}{N} \sum_{m=-(N-1)}^{N-1} \left(1 - \frac{|m|}{N}\right) r_{z_a z_b}^2(m). \quad (43)$$

⁵The reason for the approximation in (43) is that this expression is derived assuming that $z_a(n)$ and $z_b(n)$ are jointly Gaussian; however, the result applies to other distributions as well [30].

Hence, if $r_{z_a z_b}^2(m)$ is square-summable, $\text{var}(\hat{r}_{z_a z_b}^N(0)) = \text{var}(\hat{u}_i)$ approaches zero as $N \rightarrow \infty$. Bound (41) by

$$\text{var}(\hat{v}_l) \leq \sum_i \sum_j |\Psi_{l,j}^\dagger| |\mathbf{K}_{i,j}| |\Psi_{l,i}^\dagger|. \quad (44)$$

Then from (42) and (43), we conclude that $\text{var}(\hat{v}_l)$, and thus $\text{var}(\hat{\mathbf{v}}_{NC})$, approach zero as $N \rightarrow \infty$.

The noncompressive estimator $\hat{\mathbf{v}}_{NC}$ is thus statistically consistent since it is asymptotically unbiased and since its variance decreases to zero as $N \rightarrow \infty$. Incidentally, if $x(t)$ is a zero-mean Gaussian WSS random process, $\hat{\mathbf{v}}_{NC}$ is also asymptotically efficient [43, p. 471], meaning that the variance of the limit distribution of $\hat{\mathbf{v}}_{NC}$ achieves the Cramér-Rao bound. As we show in the Appendix, asymptotic efficiency follows from the fact that $\hat{\mathbf{u}}$ is a maximum likelihood estimate of \mathbf{u} .

3.7 Consistency of the compressive estimator

A detailed analysis of the compressive estimator's consistency is beyond the scope of this paper because it depends on the specific CS recovery algorithm is employed. Nevertheless, compressive estimates are consistent in the sense that ideally sparse vectors $\hat{\mathbf{v}}$ can be *exactly* recovered by some CS recovery algorithms, and thus the consistency of $\hat{\mathbf{u}}$ (as shown above) implies the consistency of $\hat{\mathbf{v}}_C$.

4 Examples

4.1 Estimating MA spectra with lines

Consider sampling a (nonsparse) moving average power spectrum with spectral lines using a 50 channel MC sampler ($q = 50, L = 64$). The process is generated by passing white noise (with unit variance) through a filter with transfer function $H(z) = (1 - z^{-1})(1 + z^{-1})^3$ and then adding the two sinusoidal components, $2 \cos(\frac{8\pi k}{17})$ and $2 \cos(\frac{11\pi k}{20})$. Fig. 6(a) shows the true PSD. Fig. 6(b) and Fig. 6(c) show two resulting noncompressive estimates for different values of N , the number of MC samples per channel. For comparison, a coarse resolution approximation of the true PSD is overlaid on the estimates. Because $\hat{\mathbf{v}}_{NC}$ is consistent, its mean squared error with respect to this coarse resolution approximation decreases to zero as $N \rightarrow \infty$. This is depicted in Fig. 6(c). Fig. 7 further illustrates the consistency of $\hat{\mathbf{v}}_{NC}$ within the context of this example for various (L, q) values. In each case, the average squared error monotonically decreases as N increases.

4.2 Sparse multiband power spectra

Suppose a real WSS random process $x(t)$ has a sparse multiband power spectrum band-limited to 1 GHz ($W = 2$ GHz). Suppose also that we are interested in a PSD estimate having a resolution of about 15 MHz. To satisfy this requirement we choose $L = 128$ which gives a resolution of 15.625 MHz. Suppose also that the sparse PSD is composed of two active bands, each with an approximate bandwidth of 30 MHz. Being conservative, we thus expect \mathbf{v} to be a 16-sparse vector (including positive and negative frequencies with each active band taking up 4 adjacent subbands).

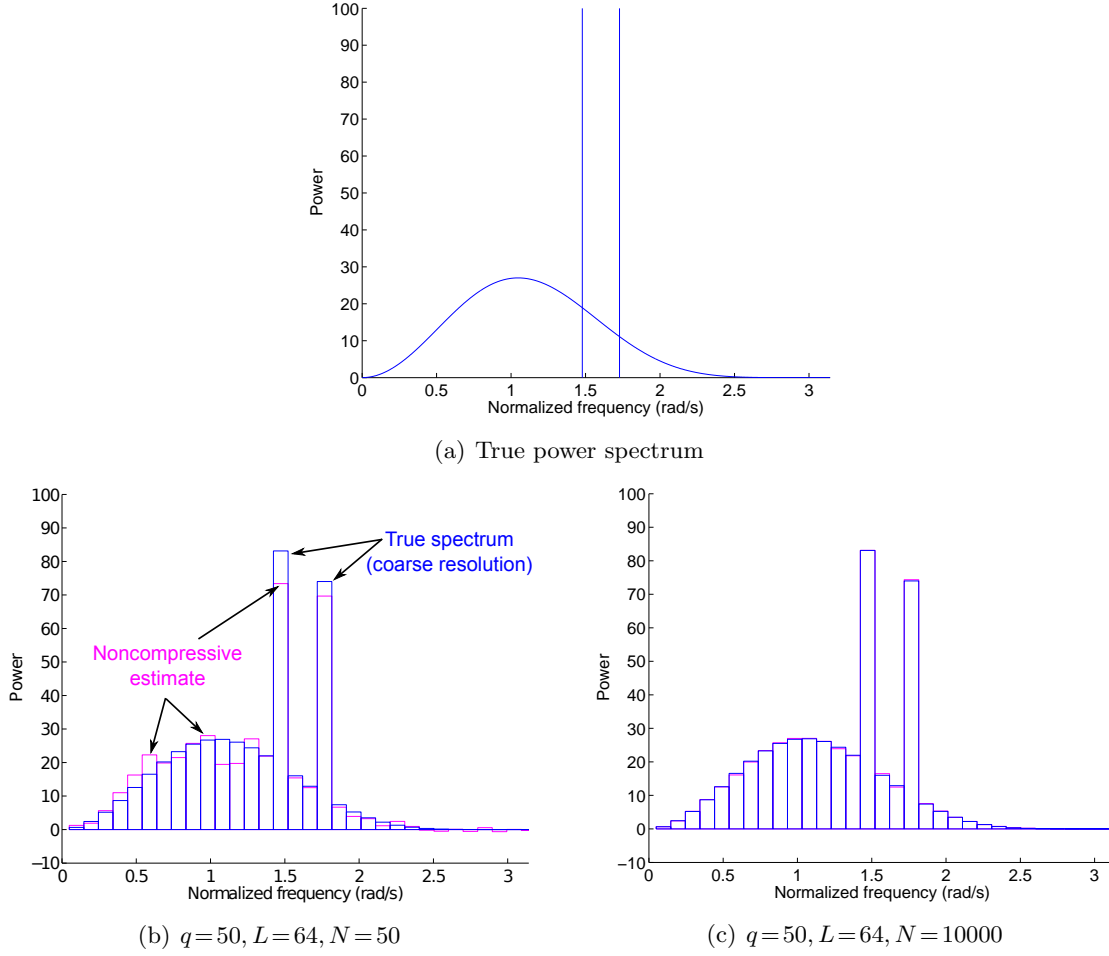


Figure 6: Estimating MA spectra with lines. (a) True power spectrum. (b) Overlay plot of a noncompressive estimate and a coarse resolution approximation of the true power spectrum. The height of the bars represent the average power in the band which it spans. (c) Estimate of same spectrum except the number of samples per channel, N , increased from 50 to 10000. Note the convergence of the noncompressive estimate.

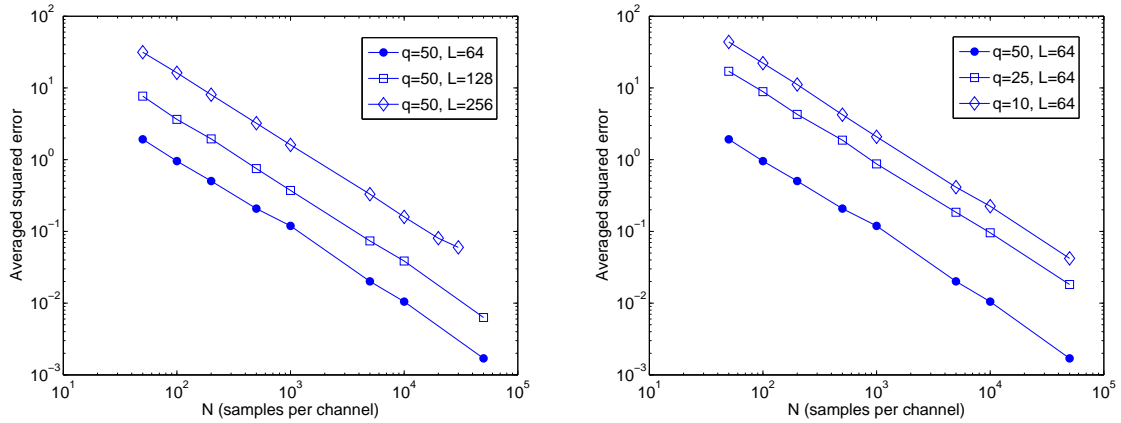


Figure 7: Decrease in average squared error for various noncompressive scenarios as $N \rightarrow \infty$.

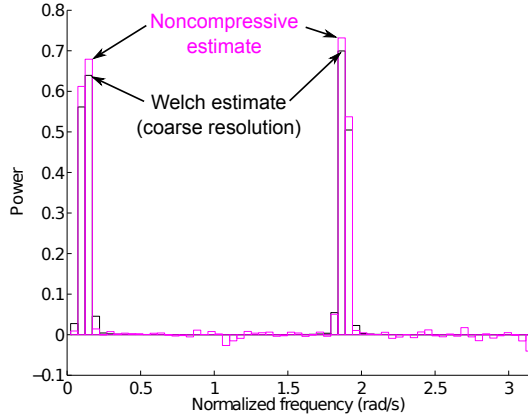
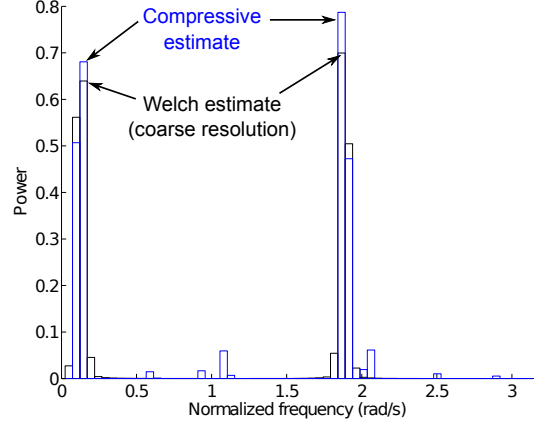
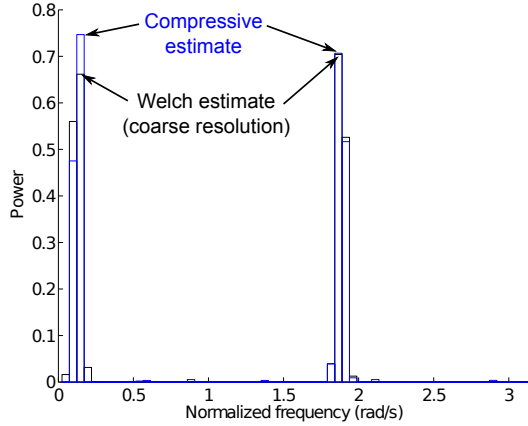
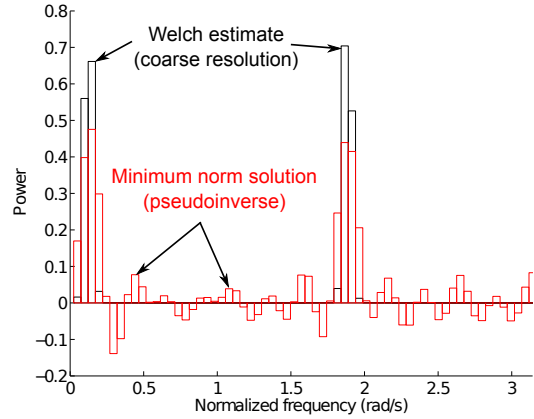
(a) $q=20, L=128, N=1000$ (overdetermined)(b) $q=7, L=128, N=1000$ (underdetermined)(c) $q=7, L=128, N=10000$ (underdetermined)(d) $q=7, L=128, N=10000$ (underdetermined)

Figure 8: Estimating sparse multiband power spectra. The overlay plots compare compressive and noncompressive estimates to a coarse resolution Welch estimate (black) based on Nyquist rate samples. (a) Noncompressive estimate (magenta) with random sampling pattern. (b)-(c) Compressive estimates (blue) with Golomb ruler sampling pattern. (d) Minimum norm least squares estimate (red) in underdetermined case.

The PSD can be estimated in either a noncompressive or a compressive manner. Choosing $q = 20$, leads to an overdetermined system in (21) ($q(q-1) + 1 = 381 \geq 128 = L$) and a noncompressive estimate. Fig. 8(a) shows the resulting noncompressive estimate with a normalized square error of 0.010. The error is calculated with respect to a coarse Welch estimate, i.e., it has the same resolution as the noncompressive estimate. The Welch estimate is based on uniform Nyquist rate samples and is computed with no overlap. The sampling pattern was chosen uniformly at random from the set $\{0, \dots, 127\}$ and it was verified that $\tilde{\Psi}$ had full rank. The fractional delays were implemented by interpolating the MC sequences to the Nyquist rate, shifting them by the appropriate delays, and then subsampling them to the original rate. Each channel collected 1000 MC samples and the average system sampling rate was about 1/6 of the Nyquist rate.

According to Corollary 1, the number of measurements required to recover a 16-sparse vector needs to be greater than or equal to 32. This implies that q needs to be at least 7 because the number of measurements in (28) equals $q(q-1) + 1$ and for $q < 7$ this quantity is less than 32. In addition, Corollary 1 requires that the sampling pattern produces the set of differences $0, \dots, 15$ for a 16-sparse signal. Fortunately, a Golomb ruler of order 7 ($\{c_i\} = \{1, 3, 4, 11, 17, 22, 26\}$) exists that produces the necessary difference set. Fig. 8(b) shows the resultant compressive estimate (normalized squared error 0.023). The reduction in the number of channels from 13 to 7, in going from the noncompressive to the compressive estimator, reduces the complexity of the sampler, or equivalently, halves the average system sampling rate.

Fig. 8(c) displays the same compressive estimate as in Fig. 8(b) except that it is computed with 10000 samples per channel instead of 1000. The normalized squared error is 0.013. In contrast to this figure, Fig. 8(d) shows the minimum (ℓ_2) norm least squares solution computed by the pseudoinverse. As one would expect, this solution is far from the true PSD and does not improve with more samples.

4.3 Spectral sensing for cognitive radio

To opportunistically use their resources, cognitive radios must be able to dynamically sense underutilized portions of the radio spectrum. Towards this end, methods and algorithms have recently been proposed to sense the largest possible bandwidth while sampling at the lowest possible rate [19]. Some of these methods have taken advantage of CS, but the following example shows that a noncompressive estimator can monitor large bandwidths at low sampling rates (although, as explained above, CS will allow better tradeoffs with a compressive estimator).

This example is only a caricature of the actual problem that must be solved to realize a practical cognitive radio system. Let $x(t)$ be a MA random process generated from filtering white Gaussian noise with a notch filter. The filter notches the spectrum such that it has approximately two 80 MHz stop bands within an overall band of 1 GHz ($W = 2$ GHz). Fig. 9 shows two noncompressive estimates and compares them to a Welch estimate of the same resolution and calculated using Nyquist samples. The left plot sets $q = 25$ and $L = 64$ and the right plot sets $q = 50$ and $L = 128$, meaning that the resolution of left hand estimate is half the resolution of the right hand estimate (half as fine) and that the sampling rate per channel on the left is twice that of the right (31.25 MHz vs 15.625 MHz). In each case, however, the overall average system sampling rate is the same (781.25 MHz). The noncompressive estimate on the left is formed with 4096 MC samples per channel, the one on the right with 2096 samples. Both plots suggest that spectral holes could be detected simply by thresholding the estimates

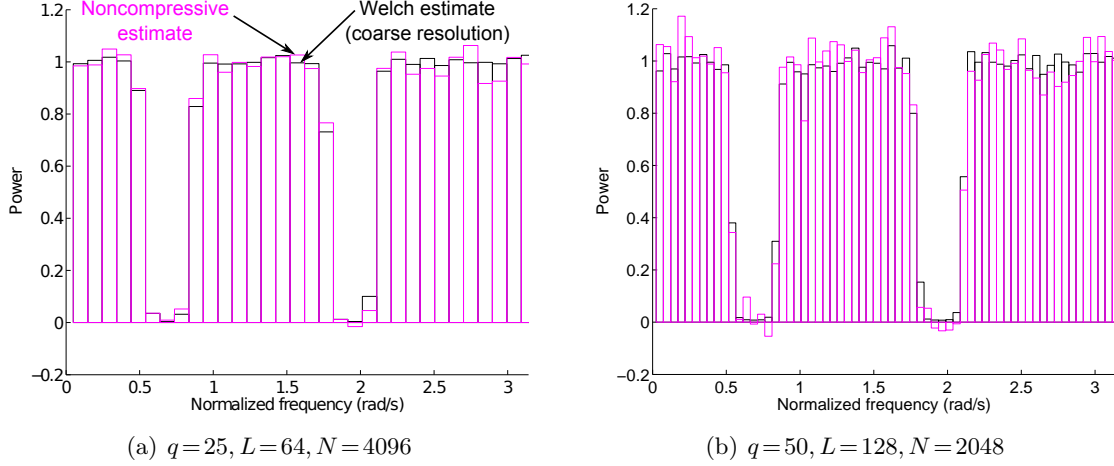


Figure 9: Estimating spectra with holes. The plots show two different noncompressive estimates (differing resolutions) of a MA spectrum with two spectral notches (holes).

and hence noncompressive estimators may provide a way to monitor large bandwidths at low sampling rates.

5 Conclusion

In this paper, we derived and analyzed a consistent, MC based PSD estimator that produces both compressive and noncompressive estimates at sub-Nyquist sampling rates. The estimator does not estimate the power spectrum on a discrete grid of frequencies but instead computes average power within a given set of subbands. Compressive estimates leverage the sparsity of the spectrum and exhibit better tradeoffs among the estimator's resolution, complexity, and average sampling rate compared to noncompressive estimates. Given suitable sampling patterns, both compressive and noncompressive estimates can be recovered using NNLS, thus avoiding the overhead of computing the estimates in different ways. The estimator is consistent and has wide applicability; it is especially attractive in wideband applications where high sampling rates are costly or difficult to implement.

6 Appendix

To show that $\hat{\mathbf{v}}_{NC}$ is asymptotically efficient, we show that it is a maximum likelihood estimator (MLE) because MLEs are known to be asymptotically efficient [43, p. 472].

Define $\mathbf{z}(n)$ to be the n th snapshot of samples across all the channels of a MC sampler:

$$\mathbf{z}(n) \triangleq [z_1(n), z_2(n), \dots, z_q(n)]^T. \quad (45)$$

The entries of $\mathbf{z}(n)$ are the nonuniform samples collected within one period (L/W seconds) of the MC sampling sequence. The upper triangular portion of the sample correlation matrix \mathbf{S}

are the elements of $\hat{\mathbf{u}}$:

$$\mathbf{S} \triangleq \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{z}(n) \mathbf{z}(n)^T \quad (46)$$

$$= \begin{bmatrix} \hat{r}_{z_1 z_1}^N(0) & \hat{r}_{z_1 z_2}^N(0) & \dots & \hat{r}_{z_1 z_q}^N(0) \\ \hat{r}_{z_2 z_1}^N(0) & \hat{r}_{z_2 z_2}^N(0) & \dots & \vdots \\ \vdots & & \ddots & \\ \hat{r}_{z_q z_1}^N(0) & \dots & \dots & \hat{r}_{z_q z_q}^N(0) \end{bmatrix}. \quad (47)$$

Suppose $x(t)$ is zero-mean WSS Gaussian random process. Then each snapshot $\mathbf{z}(n)$ is an i.i.d. realization of a jointly Gaussian random vector with correlation matrix \mathbf{R} . Taking \mathbf{R} to be the parameter of interest, the log-likelihood function given the data $\{\mathbf{z}(n)\}_{n=0}^{N-1}$ is [44, 45]

$$L(\mathbf{R}; \{\mathbf{z}(n)\}) = \frac{-qN}{2} \ln 2\pi - \frac{N}{2} \ln \det(\mathbf{R}) - \frac{N}{2} \text{tr}[\mathbf{R}^{-1} \mathbf{S}]. \quad (48)$$

The maximum likelihood estimate of \mathbf{R} can be found by taking the gradient of $L(\mathbf{R}; \{\mathbf{z}(n)\})$ with respect to \mathbf{R} and setting the result equal to zero.

$$\frac{\partial L(\mathbf{R}; \{\mathbf{z}(n)\})}{\partial \mathbf{R}} = -\frac{N}{2}(\mathbf{R}^{-1})^T + \frac{N}{2}(\mathbf{R}^{-1} \mathbf{S} \mathbf{R}^{-1})^T = 0 \quad (49)$$

Solving for \mathbf{R} reveals that the maximum likelihood estimate of \mathbf{R} equals the sample correlation \mathbf{S} , or in other words, the entries in $\hat{\mathbf{u}}$ are maximum likelihood estimates of $r_{z_a z_b}(0)$ for all combinations of pairs (a, b) .

Now since $\hat{\mathbf{v}}_{NC} = \Psi^\dagger \hat{\mathbf{u}}$, we have by the invariance property of maximum likelihood estimators [43, 45] that $\hat{\mathbf{v}}_{NC}$ is a maximum likelihood estimate of \mathbf{v} . We therefore conclude that $\hat{\mathbf{v}}_{NC}$ is asymptotically efficient when $x(t)$ is a zero-mean WSS Gaussian random process.

References

- [1] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Th.*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [2] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Info. Th.*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [3] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Th.*, vol. 52, no. 4, pp. 1289–1306, Apr 2006.
- [4] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 21–30, Mar 2008.
- [5] J. Romberg, "Compressive sensing by random convolution," *SIAM J. Imaging Sci.*, vol. 2, no. 4, pp. 1098–1128, 2009.
- [6] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Info. Th.*, vol. 56, no. 1, pp. 520–544, 2010.

- [7] M. Mishali and Y.C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Sig. Process.*, vol. 4, no. 2, pp. 375–391, 2010.
- [8] M. A. Lexa, M. E. Davies, and J. S. Thompson, "Reconciling compressive sampling systems for spectrally sparse continuous-time signals," *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 155–171, Jan. 2012.
- [9] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," *Proc. IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1688–1691, May 1996.
- [10] Ping Feng, *Universal Minimum-Rate Sampling and Spectrum-Blind Reconstruction for Multiband Signals*, Ph.D., University of Illinois at Urbana-Champaign, 1997.
- [11] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Trans. Info. Th.*, vol. 46, no. 6, pp. 2173–2183, Sep 2000.
- [12] Y. Bresler, "Spectrum-blind sampling and compressive sensing for continuous-index signals," *IEEE Info. Th. and Appl. Workshop*, pp. 547–554, Jan 2008.
- [13] M. Mishali and Y. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 993–1009, 2009.
- [14] Cormac Herley and Ping Wah Wong, "Minimum rate sampling and reconstruction of signals with arbitrary frequency support," *IEEE Trans. Info. Th.*, vol. 45, no. 5, pp. 1555–1564, Jul. 1999.
- [15] P.P. Vaidyanathan and Vincent C. Liu, "Efficient reconstruction of band-limited sequences from nonuniformly decimated versions by use of polyphase filter banks," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 38, no. 11, pp. 1927–1936, Nov. 1990.
- [16] M.A. Lexa, M.E. Davies, and J.S. Thompson, "Compressive spectral estimation can lead to improved resolution/complexity tradeoffs," in *4th Workshop on Signal Processing with Adaptive Sparse Structured Representations*, Jun 2011, p. 72.
- [17] Z. Tian and G.B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing*, 2007, vol. 4, pp. 1357–1360.
- [18] Geert Leus and Dyonisius Dony Ariananda, "Power spectrum blind sampling," *IEEE Signal Processing Letters*, pp. 443–450, Aug. 2011.
- [19] Dyonisius Dony Ariananda and Geert Leus, "Wideband power spectrum sensing using sub-Nyquist sampling," in *Proc. IEEE Inter. Workshop on Sig. Processing Advances in Wireless Comm.*, Jun 2011.
- [20] D.D. Ariananda, G. Leus, and Z. Tian, "Multi-coset sampling for power spectrum blind sensing," in *Digital Signal Processing (DSP), 2011 17th International Conference on*, Jul. 2011, pp. 1–8.

- [21] Petre Stoica, Prabhu Babu, and Jian Li, “New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data,” *IEEE Trans. Signal Processing*, vol. 59, no. 1, pp. 35–47, 2011.
- [22] Petre Stoica, Prabhu Babu, and Jian Li, “SPICE: A sparse covariance-based estimation method for array processing,” *IEEE Trans. Signal Processing*, vol. 59, no. 2, pp. 629–638, 2011.
- [23] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, New York, 3 edition, 1991.
- [24] Richard Roberts and Clifford Mullis, *Digital Signal Processing*, Addison-Wesley Publishing Co., Inc., 1987.
- [25] T.I. Laakso, V. Valimaki, M. Karjalainen, and U.K. Laine, “Splitting the unit delay: Tools for fractional delay filter design,” *IEEE Signal Processing Mag.*, vol. 13, no. 1, pp. 30–60, Jan 1996.
- [26] E. Cande’s, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [27] D. Needell and J. Tropp, “COSAMP: Iterative signal recovery from incomplete and inaccurate samples,” *Appl. Computational Harm. Anal.*, vol. 26, pp. 301–321, 2008.
- [28] T. Blumensath and M. Davies, “Iterative hard thresholding for compressed sensing,” *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.
- [29] Mark Rudelson and Roman Vershynin, “On sparse reconstruction from Fourier and Gaussian measurements,” *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, Aug. 2008.
- [30] D.H. Johnson and D.E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice Hall, Upper Saddle River, NJ, 1993.
- [31] Athanasios Papoulis, *The Fourier Integral and Its Applications*, McGraw-Hill, Inc., 1962.
- [32] Konstantinos Drakakis, “A review of the available construction methods for Golomb rulers,” *Advances in Mathematics of Communications*, vol. 3, no. 3, pp. 235–250, Aug. 2009.
- [33] David L. Donoho and Jared Tanner, “Counting the faces of randomly-projected hypercubes and orthants, with applications,” *Discrete and Computational Geometry*, vol. 43, no. 3, pp. 522–541, 2010.
- [34] David L. Donoho and Jared Tanner, “Counting the faces of randomly-projected hypercubes and orthants, with applications,” [Online] arXiv:0807.3590v1 [math.MG], 2008.
- [35] A.M. Bruckstein, M. Elad, and M. Zibulevsky, “On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations,” *IEEE Trans. Info. Th.*, vol. 54, no. 11, pp. 4813–4820, 2008.

- [36] Martin Slawski and Matthias Hein, “Robust sparse recovery with non-negativity constraints,” in *4th Workshop on Signal Processing with Adaptive Sparse Structured Representations*, Jun 2011, p. 30.
- [37] David L. Donoho and Jared Tanner, “Sparse nonnegative solution of underdetermined linear equations by linear programming,” *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 27, pp. 9446–9451, 2005.
- [38] Bashar I. Ahmad and Andrzej Tarczynski, “Reliable wideband multichannel spectrum sensing using randomized sampling schemes,” *Signal Processing*, vol. 90, pp. 2232–2242, 2010.
- [39] Andrzej Tarczynski and Dongdong Qu, “Optimal random sampling for spectrum estimation in DASP applications,” *Int. J. Appl. Math. Comput. Sci.*, vol. 15, no. 4, pp. 463–469, 2005.
- [40] E. Masry, “Poisson sampling and spectral estimation of continuous-time processes,” *IEEE Trans. Info. Th.*, vol. 24, no. 2, pp. 173 – 183, Mar 1978.
- [41] E. Masry, “Alias-free sampling: An alternative conceptualization and its applications,” *IEEE Trans. Info. Th.*, vol. 24, no. 3, pp. 317–324, May 1978.
- [42] F. Beutler, “Alias-free randomly timed sampling of stochastic processes,” *Information Theory, IEEE Transactions on*, vol. 16, no. 2, pp. 147 – 152, Mar 1970.
- [43] George Casella and Roger L. Berger, *Statistical Inference*, Duxbury, 2002.
- [44] C.T. Mullis and Louis L. Scharf, *Modern Spectral Estimation*, chapter Quadratic Estimators of the Power Spectrum, pp. 1–57, Houghton-Mifflin, 1990.
- [45] L.L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley Publishing Co., Reading, MA, 1991.